

28. Morishita, W., Kirew, S. A. & Alpert, B. E. Evidence for metabotropic glutamate receptor activation in the induction of depolarization-induced suppression of inhibition in hippocampal CA1. *J. Neurosci.* 18, 4870–4882 (1998).
29. Marr, D. A theory of cerebellar cortex. *J. Physiol. (Lond.)* 202, 437–470 (1969).
30. Ito, M. In *The Cerebellum and Neural Control* (Raven, New York, 1984).

Acknowledgements

This work was supported by The Wellcome Trust, the European Union and the Medical Research Council (Research Studentship to S.J.M.). We thank A. Momiyama and T. Takahashi for helpful discussion, J. Clements for providing Axograph and D. Attwell, M. Parment, M. Häusser, B. Katz and T. Takahashi for comments on the manuscript.

Correspondence and requests for materials should be addressed to R.A.S. (e-mail: a.silver@ucl.ac.uk).

Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491

J. Parkhill*, M. Achtman†, K. D. James*, S. D. Bentley*, C. Churcher*, S. R. Klee†, G. Morelli†, D. Basham*, D. Brown*, T. Chillingworth*, R. M. Davies*, P. Davis*, K. Devlin*, T. Feltwell*, N. Hamlin*, S. Holroyd*, K. Jagels*, S. Leather*, S. Moutle*, K. Mungall*, M. A. Quail*, M.-A. Rajandream*, K. M. Rutherford*, M. Simmonds*, J. Skelton*, S. Whitehead*, B. G. Spratt‡ & B. G. Barrell*

* The Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

† Max-Planck Institut für molekulare Genetik, Ihmestraße 73, D-14195 Berlin, Germany

‡ Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3FY, UK

Neisseria meningitidis causes bacterial meningitis and is therefore responsible for considerable morbidity and mortality in both the developed and the developing world. Meningococci are opportunistic pathogens that colonize the nasopharynx and oropharynx of asymptomatic carriers. For reasons that are still mostly unknown, they occasionally gain access to the blood, and subsequently to the cerebrospinal fluid, to cause septicaemia and meningitis. *N. meningitidis* strains are divided into a number of serogroups on the basis of the immunochemistry of their capsular polysaccharides; serogroup A strains are responsible for major epidemics and pandemics of meningococcal disease, and therefore most of the morbidity and mortality associated with this disease. Here we have determined the complete genome sequence of a serogroup A strain of *Neisseria meningitidis*, Z2491 (ref. 1). The sequence is 2,184,406 base pairs in length, with an overall G+C content of 51.8%, and contains 2,121 predicted coding sequences. The most notable feature of the genome is the presence of many hundreds of repetitive elements, ranging from short repeats, positioned either singly or in large multiple arrays, to insertion sequences and gene duplications of one kilobase or more. Many of these repeats appear to be involved in genome fluidity and antigenic variation in this important human pathogen.

The genome encodes complete sets of enzymes for glycolysis (apart from *fruK* and *pfkA*), gluconeogenesis, the pentose-phosphate and Entner–Doudoroff pathways, the pyruvate dehydrogenase complex and the trichloroacetic acid cycle. In addition to the aerobic respiration genes, there are also both nitrite (*aniA*) and nitrate (NMA1886) reductases; some capability for fermentation is

also apparent. *N. meningitidis* appears to be capable of *de novo* synthesis of most of the amino acids (with the exception of asparagine and methionine) and purine and pyrimidine nucleotides; however, the pathways for production of folic acid, molybdopterin, pantothenate and pyridoxine are incomplete. All of the aminoacyl transfer RNA synthetases are present except tRNA^{Asn}. As *N. meningitidis* encodes both glutamyl tRNA synthetase and the three-subunit Glu-tRNA^{Gln} transamidase, the latter may be responsible for the production of tRNA^{Asn} by the transamidation of tRNA^{Sp}.

In addition to the 2,121 predicted coding sequences (CDSs), there are 4 copies of a 16S–23S–5S ribosomal RNA operon, 58 tRNAs, 1 tmRNA (10Sa RNA) and the RNA component of RNAase P. The average gene length is 877 base pairs (bp), at a density of 1 per 1.03 kilobases (kb). The overall coding density is 82.9%, higher than *Rickettsia*², but lower than most other sequenced bacteria. Base 1 of the sequence was chosen to correspond roughly with 12 o'clock on the published map³; the origin of replication (as indicated by the bias towards G on the leading strand⁴) is around 247,600 bp. Intriguingly, although the GC bias is similar in both replicons, there is a clear bias (60.2%) of CDSs towards the leading strand in only one replicon (anticlockwise in Fig. 1), with no coding bias at all on the opposite replicon. The genome contains at least 56 pseudogenes, of which 17 are remnants of insertion sequence (IS) elements, and there are at least 5 complete or partial prophages (pnm1–5).

The G+C content of the genome is extremely variable, with at least 60 coding regions (nearly 5% in total) having a significantly lower G+C content, ranging in size from 224 bp to 11.3 kb and averaging 1.8 kb (see Supplementary Information). These regions may represent recently acquired DNA, and given the natural competence of *Neisseria* species⁵, this is perhaps unsurprising. Although most genes in these regions have no database matches, or are conserved hypothetical proteins, nearly one-third encode

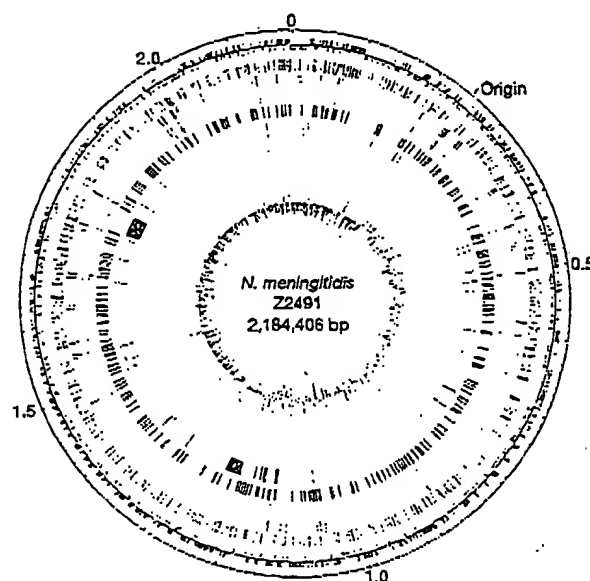


Figure 1 Circular representation of the *N. meningitidis* Z2491 genome. The concentric circles show, reading inwards: the scale in megabases, with the origin of replication indicated; predicted coding sequences clockwise (dark green) and anti-clockwise (light green); neisserial uptake sequences (red); dPS3 sequences (dark orange); RS elements (light orange); dispersed repeats (Correlia, ATR, REP2-5; black); IS elements and phage (narrow ticks and wide bars respectively; turquoise) and tandem repeats (dark blue). The inner histogram shows plot of (G-C)/(G+C) with values greater than zero in yellow and less than zero in orange. Figure generated with LASERGENE software (DNASTAR).

proteins that are likely to be on the surface of the cell, or are responsible for the production of surface structures. Prominent among these are the well characterized CDSs of the serogroup A capsule cassette (*sacA-D*) which produce a serogroup-A-specific sugar residue in the capsular polysaccharide⁶; these are located in an extremely well defined region of only 29.7% G+C. Also notable are a single glycosyl-transferase (NMA0832), similar to capsule biosynthesis genes in other bacteria, and a pilin homologue (NMA0424). Some surface-exposed protein-encoding genes are each immediately followed by large regions of A+T-rich DNA encoding numerous small hypothetical proteins; these include the three *mafB* family genes and NMA1083. Eight A+T-rich regions encode DNA restriction/modification enzymes. Some caution is required in interpreting a bias in G+C composition as necessarily indicating recently acquired DNA; the largest region of A+T-rich DNA (41.7% G+C over 11.3 kb) encompasses much of the ribosomal protein operon.

One of the most striking and unique characteristics of the Z2491 genome is the abundance and diversity of repetitive DNA (Fig. 1). The most obvious example of this abundance is the Neisserial uptake sequence, which is involved in the recognition and uptake of DNA from the environment⁷. There are nearly two thousand copies of the 10-bp uptake sequence, which either occurs alone or in inverted repeats as part of a transcriptional terminator⁷. Other repetitive sequence elements in the *N. meningitidis* genome are concentrated within intergenic repeat arrays of 200–2,700 bp. These repeat arrays are composed of several different repeat types (Fig. 2 and Table 1), the most abundant of which are the 'neisserial intergenic mosaic elements' (NIMEs), which comprise repeat units of ~50–150 bp (RS elements), each flanked by 20-bp inverted repeats (dRS3 elements); there are between 1 and 60 copies of 117 different families of RS element.

Also present in the repeat arrays are larger units, which are also found in isolation, including the 'Correia elements' (CEs, 156-bp sequences bounded by 26-bp inverted repeats) that have been previously described in both *N. meningitidis* and *Neisseria gonorrhoeae*⁸. Almost one-third of the 257 CEs share a common 51-bp internal deletion, which presumably has not affected the ability of the element to be mobilized, and 29 have lost one or both

of their terminal inverted repeats. Correia elements in repeat arrays are sometimes flanked by additional conserved sequences upstream and downstream (Fig. 2). A number of both complete and internally deleted CEs occur embedded within other sequence features, including genes, RS elements, other CEs and ISs. The pseudogenes NMA0059 (a modification methylase) and NMA0530 (a hypothetical protein) both appear to have been truncated or interrupted by CE insertion. Two repeat types are dispersed over the genome at low frequency, both alone and within repeat arrays: 19 copies of a 183-bp A+T-rich (30% G+C) repeat (ATR) whose ends form an imperfect 35-bp inverted repeat and 26 copies of a 120–150-bp repeat (REP 2). REP 2 occurs immediately upstream of 16 CDSs and contains a ribosome-binding-site-like conserved AAGGA motif within 5–13 bp of the predicted start codon. It seems likely that sequences within this repeat would exert some effect on the expression of these genes. The repeat arrays appear to be a target for the insertion of the mobile elements in the genome, with almost half of the 43 complete and partial copies of IS1016, IS1106 and IS1655 being integrated in or near to a repeat array.

Given the natural competence of neisserial species, the presence of such arrays could encourage sequence variation by acting as a target for a specific recombinase, thereby increasing the rate of horizontal gene transfer at the flanking locus, resulting in rapid allelic replacement or the creation of new mosaic alleles. A previously characterized example of enhanced recombinational exchange involves the rearrangements responsible for the modulation of *pilE* expression. *PilE* is the main structural pilin component and is expressed from the intact *pilE* gene. In *N. gonorrhoeae*, 'silent' pilin genes (*pilS*) exist, which contain only the 3' region of the gene without a start codon or transcriptional signals⁹. Coding sequences from these silent regions can be recombined into the expression locus, in a *recA/Q/O*-dependent manner, changing the expressed *PilE* sequence¹⁰. Models put forward for this process suggest that the gene conversion may be intergenomic¹¹. In the *N. meningitidis* Z2491 sequence, the *pilE* gene is surrounded by repeat arrays consisting of NIMEs and a CE (Fig. 3a). Immediately upstream of the *pilE* gene are eight *pilS* loci, each without an 5' end and intimately bounded by arrays of NIMEs, suggesting that NIME sequences may be involved in *pilS/E* recombination. The *N. gonorrhoeae* *Sma*/Cln repeat, implicated in enhancing *pilE/S* recombination in *N. gonorrhoeae*¹², exists as a single copy in *N. meningitidis* Z2491, immediately downstream of the *pilE* gene.

Our analysis of the occurrence of repeat arrays suggests that antigenic variation mediated by gene conversion, similar to that of *pilE*, may occur widely in the *N. meningitidis* genome. In *N. meningitidis* Z2491, most shorter repeat arrays are associated with genes directly or indirectly related to cell-surface functions (including pilins, substrate-binding proteins and transporter components), and the larger repeat arrays are exclusively associated with such genes (Fig. 4; and Table 2, Supplementary Information). The

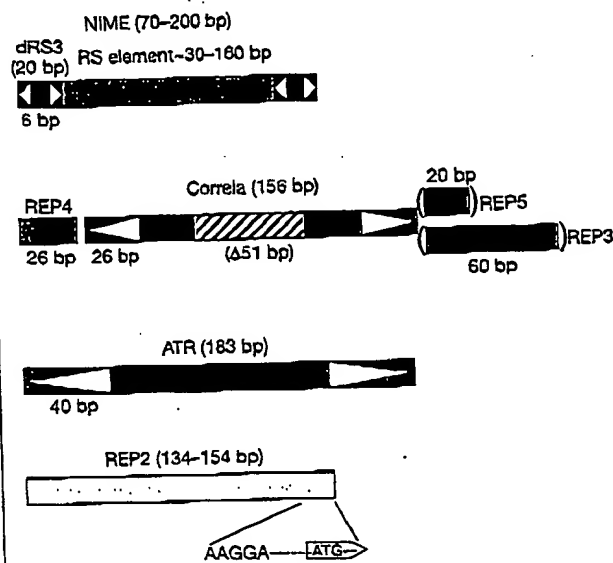


Figure 2 Types of *N. meningitidis* repeat. The name of each repeat type is indicated above the repeat. Inverted repeat sequences are represented by open triangles, and the internal deletion present in some Correia elements by a hatched box. The translational signals within REP2 are indicated below the repeat.

Table 1 Number of repeats by type in *N. meningitidis* Z2491

Type	Size (bp)	Frequency
DNA uptake sequence: ggcgcctgag	10	1,852
RS	24–161	681
dRS3: attccnnnnnnnnngggatt	20	772
Correia (full)	150–159	173
Correia (internal deletion)	~104	84
Correia (partial)	37–145	29
ATR	183	19
REP 2	59–164	26
REP 3	60	13
REP 4	26	20
REP 5	20	9
IS1016	256–740	14 (including partial)
IS1106	203–1219	22 (including partial)
IS1655	1,074–1,257	7 (including partial)
Prophage	2,330–38,954	5

products of these genes are potential targets for the immune system, and antigenic variation of these proteins would be beneficial in evading the immune response. The genes encoding the outer membrane proteins PorA (NMA1642) and PorB (NMA0398) are flanked on both sides by arrays of repeats, as are *lbpAB* (ref. 13) (Fig. 3b), *tbpAB* (ref. 14) and *hpuAB* (ref. 15), which encode lactoferrin-, transferrin- and haemoglobin-/haptoglobin-binding proteins, respectively. The putative ferric enterobactin-binding protein genes *setAB* are also associated with two such arrays. In contrast, the neisserial ferric binding protein operon *fbpABC* (ref. 16), which may not be exposed to the immune response, the TonB-dependent receptor NMA0577 and the haemoglobin receptor *hmbR* (ref. 17) are not associated with repeat elements (a poly-(G) tract is present in *hmbR*, potentially allowing phase-variable expression; see below). The *opaD*, *opaB*, *opaA* and *opcA* genes, which are involved in various interactions with the host cell, are all flanked on one or both sides by repeat arrays. Comparison of these regions

from a range of *N. meningitidis* isolates has shown that the repeat arrays themselves are highly variable between diverse *Neisseriae* but may be identical in related strains even after more than 30 years of epidemic spread (G.M., unpublished data).

Repeat-mediated gene rearrangements are also apparent at other sites, although, in these cases, the repeats are local. One such site encompasses the *mafA* and *mafB* genes; *mafA* is predicted to be a lipoprotein and *mafB* to be secreted; both have been proposed to have adhesin activity in *N. gonorrhoeae* (S. Eickernjaeger *et al.*, personal communication; EMBL accession number AF142582). The *N. meningitidis* orthologue of *mafB* is immediately followed by a 6.5-kb region of low G+C DNA encoding mainly short genes of unknown function. Interspersed with these genes, and with a normal G+C content, are three repeats of 303 bp and one of 79 bp from *mafB*, each corresponding to the beginning of an open reading frame (ORF) that has no start codon. This arrangement suggests that these repeat sequences may be capable of recombination

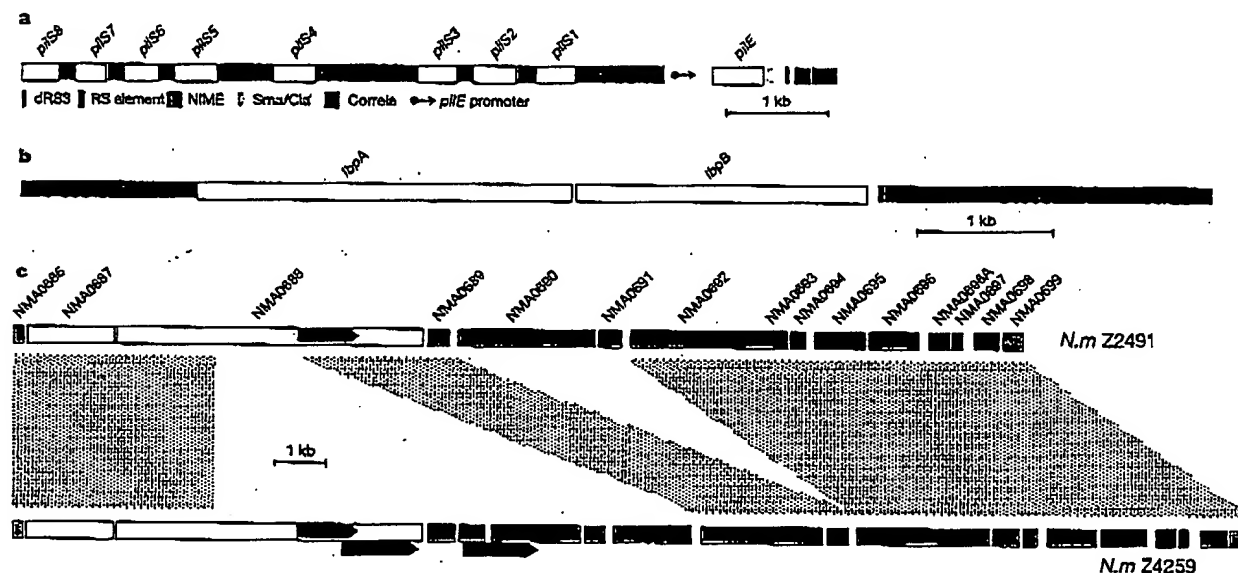


Figure 3 Structure of selected repeat regions. **a**, Repeats around the *pilE/S* locus. *pil* genes are indicated by open boxes and repeats by coloured boxes as described in the key. **b**, Large repeat arrays around the *lbpA* and *lbpB* genes. Repeats are coloured as in **a**. **c**, The filamentous haemagglutinin homologue gene sequences of *N. meningitidis* Z2491 and Z4259. Conserved sequences are indicated by light blue bars, and internal repeat

regions by red arrows. The complex repeat structures have been simplified for clarity. Open boxes represent genes encoding surface-exposed proteins, and green boxes represent genes with no database matches. The brown boxes represent the two halves of an ABC transporter pseudogene.

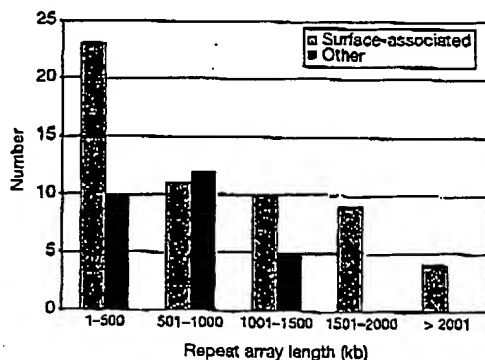


Figure 4 Graph showing the relationship between repeat array length and flanking gene function. For each category of repeat array length the number of flanking genes

associated with surface structures is shown in grey, and the number of genes in all other categories is shown hatched.

with the *mafB* gene, leading to the attachment of different 3' sequences. This interpretation is supported by the observation that the 3' sequence of the gonococcal *mafB* orthologue is almost identical to the ORF following the first 303-bp repeat, and not to the 3' sequence of the meningococcal *mafB* gene.

A second example of repeat-mediated rearrangement of the 3' ends of genes encoding surface-exposed proteins is provided by NMA0688, a homologue¹⁸ of the *Bordetella pertussis* filamentous haemagglutinin *shaB* (ref. 19). The sequence downstream of NMA0688 contains complex repeated sequences from NMA0688 (Fig. 3c) and encodes several proteins of unknown function. Sequencing this region from *N. meningitidis* Z4259 (a serogroup C strain; EMBL accession number AJ391284) revealed a similar repeat structure, but with additional DNA inserted into the 3' sequence of NMA0688 at one of the repeats. This has the effect of changing the 3' end of the NMA0688 gene, as well as introducing additional genes, some containing similar repeats. The 3' region of the Z4259 NMA0688 gene is within a downstream ORF in Z4259. It therefore seems likely that, like *mafB*, the 3' end of this *shaB* homologue can be altered by recombination with alternative 3' ends. Intriguingly, the entire region consisting of the *shaC* (NMA0687) and *shaB* homologues, along with the downstream DNA, appears to have been inserted into the coding sequence of an ABC transporter pseudogene (NMA0686/NMA0699).

Tandem repeats of nucleotides in *Neisseria* are subject to lengthening or shortening during replication owing to a phenomenon called slipped-strand mispairing²⁰. When these tandem repeats are present in coding sequences, this change in length can change the translation state of the gene, leading to an on/off switching of the gene product called 'phase variation'. The complete genome sequence reveals around 26 tandem repeats indicating potentially phase-variable genes (Table 3, Supplementary Information), many of which have been studied before. These repeats range from homopolymeric tracts of G or C nucleotides to di-, tetra- and penta-nucleotide repeats. As would be expected, most phase-variable genes are either surface-exposed or involved in the biosynthesis or modification of surface structures, which supports their suggested role in virulence and/or immune avoidance. Four of the tandem repeats are not directly within coding sequences, but are immediately upstream of candidate variable genes, and these may affect transcription from gene-specific promoters, as has been shown for *opcA* (ref. 21) and *porA* (ref. 22).

The analysis presented here supports the general operation of three mechanisms of repeat-mediated antigenic variation within the *N. meningitidis* genome: on/off switching and transcriptional modulation of gene expression by slipped-strand mispairing of short tandem repeats; intragenomic recombination of localized repeats leading to the use of different carboxy termini for surface-exposed proteins; and intergenomic gene conversion of specific surface-associated genes associated with large arrays of global repeats, mediated by the internalization of related DNA through the highly repetitive DNA uptake sequence. □

Methods

DNA was prepared from *N. meningitidis* Z4259 as described²³. The DNA was fragmented by sonication, size-fractionated on an agarose gel, and two libraries were generated in pUC18 using size fractions ranging from 0.5 to 1.5 kb. Roughly 37,500 pUC clones were sequenced from both ends using Dye-terminator chemistry on ABI 373 and 377 sequencing machines; 58,269 reads were used to generate the final assembly, giving about a 10-fold coverage of the genome. Sequence assembly was accomplished using Phrap (P. Green, unpublished), and the sequencing was finished using GAP4 (ref. 24). The assembly was verified by genomic PCR reactions across all unbridged large repeats, in addition to 260 forward and reverse reads from a random library of 20–22 kb cloned in lambdaFixII (Stratagene), and 610 forward and reverse reads from 3 libraries of 9–13 kb cloned in pSP64 (Promega). The final assembly was checked against the published map using the positions of restriction sites, mapped genes and lambda clones. In the final assembly, less than 0.07% of the genome was covered by a single clone only, and less than 0.12% was unsequenced on both strands, or with complementary sequencing chemistries. The DNA was compared with sequences in the EMBL database using BLASTN and

BLASTX²⁵. Transfer RNAs were predicted by tRNAscan-SE (ref. 26). Potential GDSs were predicted using ORPHEUS²⁷ and GLIMMER²⁸ (both trained on an initial ORF set generated by ORPHEUS), and the results were combined and checked manually. The predicted protein sequences were searched against a non-redundant protein database using WUBLASTP and FASTA. The complete six-frame translation was used to search PROSITE, and the predicted proteins were compared against the PFAM database of protein domain hidden Markov models²⁹. The results of all these analyses were assembled together using the Artemis sequence viewer (K.M.J., unpublished) and used to inform a manual annotation of the sequence and predicted proteins. Annotation was based, wherever possible, on characterized proteins or genes. Repeat sequences were identified using HMMER (S. Eddy, personal communication) and the EMBOSS program 'profit' (<http://www.sanger.ac.uk/Software/EMBOSS/>), based on sequence alignments generated with ClustalW³⁰.

Received 1 March; accepted 8 March 2000.

1. Dempsey, J. A., Wallace, A. B. & Cannon, J. G. The physical map of the chromosome of a serogroup A strain of *Neisseria meningitidis* shows complex rearrangements relative to the chromosomes of the two mapped strains of the closely related species *N. gonorrhoeae*. *J. Bacteriol.* 177, 6390–6400 (1995).
2. Curnow, A. W., Tumbula, D. L., Pelaschier, J. T., Min, B. & Soll, D. Glutaryl-tRNA (Gln) amidotransferase in *Deinococcus radiodurans* may be confined to asparagine biosynthesis. *Proc. Natl Acad. Sci. USA* 95, 12838–12843 (1998).
3. Andersson, S. G. et al. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396, 133–140 (1998).
4. Lohry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* 13, 660–665 (1996).
5. Smith, N. H., Holmes, E. C., Donovan, G. M., Carpenter, C. A. & Spratt, B. G. Networks and groups within the genus *Neisseria*: analysis of *argP*, *recA*, *rho*, and 16S rRNA sequences from human *Neisseria* species. *Mol. Biol. Evol.* 16, 773–783 (1999).
6. Swarley, J. S. et al. Characterization of the gene cassette required for biosynthesis of the (α1–6)-linked N-acetyl-N-mannosamine-1-phosphate capsule of serogroup A *Neisseria meningitidis*. *J. Bacteriol.* 180, 1533–1539 (1998).
7. Goodman, S. D. & Scoones, J. J. Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. *Proc. Natl Acad. Sci. USA* 85, 6982–6986 (1988).
8. Corrie, F. P., Inouye, S. & Inouye, M. A family of small repeated elements with some transposon-like properties in the genome of *Neisseria gonorrhoeae*. *J. Biol. Chem.* 263, 12194–12198 (1988).
9. Haas, R. & Meyer, T. F. The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: Evidence for gene conversion. *Cell* 44, 107–115 (1986).
10. Mehr, L. J. & Seifert, H. S. Differential roles of homologous recombination pathways in *Neisseria gonorrhoeae* pilin antigenic variation, DNA transformation and DNA repair. *Mol. Microbiol.* 30, 697–710 (1998).
11. Howell-Adams, E., Wainwright, L. A. & Seifert, H. S. The size and position of heterologous insertions in a silent locus differentially affect pilin recombination in *Neisseria gonorrhoeae*. *Mol. Microbiol.* 22, 509–522 (1996).
12. Wainwright, L. A., Pritchard, K. H. & Seifert, H. S. A conserved DNA sequence is required for efficient gonococcal pilin antigenic variation. *Mol. Microbiol.* 13, 75–87 (1994).
13. Pettersson, A., Frim, T., Umar, A., van der Bieren, J. & Tammesla, J. Molecular characterization of LbpA, the second lactoferrin-binding protein of *Neisseria meningitidis*. *Mol. Microbiol.* 27, 599–610 (1998).
14. Vonder Haar, R. A., Legrain, M., Kolbe, H. V. & Jacobs, E. Characterization of a highly structured domain in Tbp2 from *Neisseria meningitidis* involved in binding to human transferrin. *J. Bacteriol.* 176, 6207–6213 (1994).
15. Lewis, L. A., Gray, E., Wang, Y. P., Roe, B. A. & Dyer, D. W. Molecular characterization of hpaAB, the haemoglobin-haptoglobin-utilization operon of *Neisseria meningitidis*. *Mol. Microbiol.* 25, 737–749 (1997).
16. Adhikari, P., Berish, S. A., Nowalk, A. J., Verdeli, K. L., Morie, S. A. & Miettinen, T. A. The *fbpABC* locus of *Neisseria gonorrhoeae* functions in the periplasm-to-cytoplasm transport of iron. *J. Bacteriol.* 178, 2145–2149 (1996).
17. Stojiljkovic, I. et al. The *Neisseria meningitidis* haemoglobin receptor: its role in iron utilization and virulence. *Mol. Microbiol.* 15, 531–541 (1995).
18. Klee, S. R. et al. Molecular and biological analysis of eight genetic islands that distinguish *Neisseria meningitidis* from the closely related pathogen *Neisseria gonorrhoeae*. *Infect. Immun.* (in press).
19. Domenighini, M. et al. Genetic characterization of *Bordetella pertussis* filamentous haemagglutinin: a protein processed from an unusually large precursor. *Mol. Microbiol.* 4, 787–800 (1990).
20. Henderson, I. R., Owen, P. & Nataro, J. P. Molecular switches—the ON and OFF of bacterial phase variation. *Mol. Microbiol.* 33, 919–932 (1999).
21. Sarkari, J., Pandit, N., Moxon, E. R. & Achtman, M. Variable expression of the *Opc* outer membrane protein in *Neisseria meningitidis* is caused by size variation of a promoter containing poly-cytidine. *Mol. Microbiol.* 13, 207–217 (1994).
22. van der Ende, A. et al. Variable expression of class 1 outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the -10 and -35 regions of the promoter. *J. Bacteriol.* 177, 2473–2480 (1995).
23. Zhou, J. & Spratt, B. G. Sequence diversity within the *argP*, *fbp* and *recA* genes of natural isolates of *Neisseria meningitidis* interspecies recombination within the *argP* gene. *Mol. Microbiol.* 6, 2135–2146 (1992).
24. Bonfield, J. K., Smith, K. F. & Staden, R. A new DNA sequence assembly program. *Nucleic Acids Res.* 23, 4992–4999 (1995).
25. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410 (1990).
26. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964 (1997).
27. Frishman, D., Mironov, A., McCox, H. W. & Gelfand, M. Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. *Nucleic Acids Res.* 26, 2941–2947 (1998).
28. Salzberg, S. L., Delcher, A. L., Kasif, S. & White, O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* 26, 544–548 (1998).

letters to nature

29. Bateman, A. *et al.* *PLoS Biol.* 3:1, 215 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res.* 27, 260–262 (1999).
30. Higgins, D. G., Blewby, A. J. & Fuchs, R. CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci.* 8, 189–191 (1992).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This research was funded by The Wellcome Trust.

Correspondence and requests for materials should be addressed to J.P. (e-mail: parkhill@sanger.ac.uk). The complete sequence and annotation can be obtained from the EMBL database with the ID NMAZ2491 (accession number AL157959), and from our web pages (http://www.sanger.ac.uk/Projects/N_meningitidis).

The duration of antigen receptor signalling determines CD4⁺ versus CD8⁺ T-cell lineage fate

Koji Yasutomo*, Carolyn Doyle†, Lucio Miele‡ & Ronald N. Germain*

* Lymphocyte Biology Section, Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

† Department of Immunology, Duke University Medical Centre, Durham, North Carolina 27710, USA

‡ Cancer Immunology Program, Cardinal Bernardin Cancer Centre, Loyola University Medical Centre, Maywood, Illinois 60153, USA

Signals elicited by binding of the T-cell antigen receptor and the CD4/CD8 co-receptor to major histocompatibility complex (MHC) molecules control the generation of CD4⁺ (helper) or CD8⁺ (cytotoxic) T cells from thymic precursors that initially express both co-receptor proteins¹. These precursors have unique, clonally distributed T-cell receptors with unpredictable specificity for the self-MHC molecules involved in this differentiation process². However, the mature T cells that emerge express only the CD4 (MHC class II-binding) or CD8 (MHC class I-binding) co-receptor that complements the MHC class-specificity of the T-cell receptor. How this matching of co-receptor-defined lineage and T-cell-receptor specificity is achieved remains unknown^{1,3,4}, as does whether signalling by the T-cell receptors, co-receptors and/or general cell-fate regulators such as Notch-1 (refs 5, 6) contributes to initial lineage choice, to subsequent differentiation processes or to both. Here we show that the CD4 versus CD8 lineage fate of immature thymocytes is controlled by the co-receptor-influenced duration of initial T-cell receptor-dependent signalling. Notch-1 does not appear to be essential for this fate determination, but it is selectively required for CD8⁺ T-cell maturation after commitment directed by T-cell receptors. This indicates that the signals constraining CD4 versus CD8 lineage decisions are distinct from those that support subsequent differentiation events such as silencing of co-receptor loci.

The AND T-cell receptor (TCR) is specific for a pigeon cytochrome *c* peptide bound to the MHC class II molecule I-E^k (ref. 7). Transgenic thymocytes expressing this TCR are efficiently selected into the CD4⁺ lineage in mice expressing wild-type I-A^b MHC class II molecules. In contrast, AND TCR transgenic mice expressing mutant I-A^b molecules that are defective in interaction with CD4 but not the TCR⁸ generate CD8⁺ but not CD4⁺ mature cells⁹. To investigate how altering CD4 co-receptor binding controls the lineage fate of AND thymocytes, we took advantage of a modified two-stage reaggregate culture system¹⁰ that allows controlled deliv-

ery of MHC-dependent and independent signals to thymocytes at distinct stages of maturation (Yasutomo *et al.*, manuscript in preparation). Immature CD69^{hi}CD4⁺CD8⁺ TCR transgenic thymocytes (double positive; DP) are incubated in dispersed culture with cells expressing the desired MHC molecule ligands, to initiate selection events (culture 1). This first TCR stimulation does not lead to silencing of the CD4 or CD8 locus in dispersed culture for up to three days, but it does rapidly induce upregulation of the activation marker CD69 that characterizes thymocytes undergoing maturation *in vivo*¹¹. When the CD69^{med/hi} thymocytes arising in culture 1 after 20 h are purified and reaggregated with either MHC-positive thymic stromal cells (TSC) or MHC-negative TSC in the presence of dendritic cells of the appropriate MHC type (culture 2), mature functional T cells expressing only a single co-receptor develop over the next 60 h. Experiments using this model showed that TCR-MHC molecule interactions are required not only in culture 1 to generate the CD69^{hi} thymocytes, but also in culture 2 to generate mature T cells. Using this approach, we asked whether the first, the second or both sets of TCR/co-receptor-MHC interaction events determined the differentiation of AND TCR transgenic T cells along the CD4 versus CD8 pathways.

Thymocytes from mice expressing only the AND TCR in the absence of MHC class II molecules (AND TCR transgenic RAG-2⁺ MHC Aβ^{-/-}, referred to as AND throughout) were stimulated in culture 1 using dendritic cells from wild-type mice (WT-DC) or mutant (Mu-DC) I-A^b transgenic mice. When purified cells of the CD69^{hi}CD4⁺CD8⁺ phenotype (CD69^{hi}DP) generated by this stimulation were reaggregated with TSC from MHC class II⁺ mice, no CD4⁺ or CD8⁺ mature T cells developed (Fig. 1a, b). Inclusion of

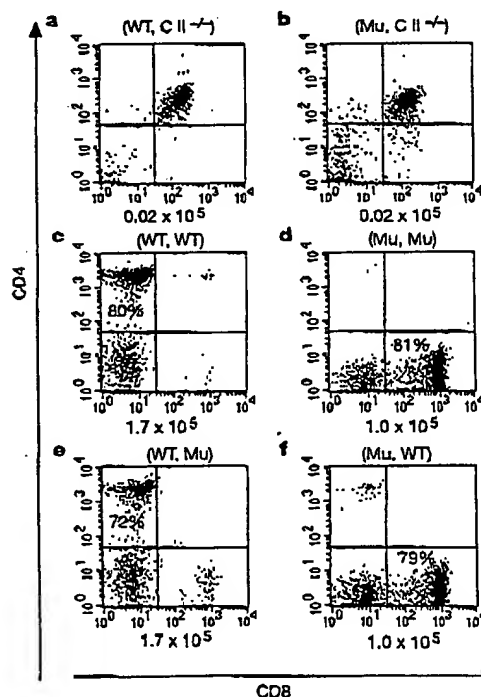


Figure 1 Role of co-receptor in CD4/CD8 lineage choice. Sorted CD69^{hi}CD4⁺CD8⁺ cells obtained after stimulation by wild-type dendritic cells (WT-DC; a, c, e) or mutant dendritic cells (Mu-DC; b, d, f) were cultured with MHC class II⁺ TSC (a, b) or MHC^{-/-} TSC and wild-type dendritic cells (WT; c, f), or mutant dendritic cells (Mu; d, e) for 72 h in thymic reagggregates. The recovered cells from all cultures were stained with anti-CD4 or CD8 monoclonal antibodies and examined by flow cytometry. The percentage of cells with a CD4⁺ or CD8⁺ phenotype is given in the upper left or lower right quadrant of each panel and the absolute number of recovered viable thymocytes is given below each panel.

Sequence analysis and expression of the aspartokinase and aspartate semialdehyde dehydrogenase operon from rifamycin SV-producing *Amycolatopsis mediterranei* ☆

Weiwen Zhang *, Weihong Jiang, Guoping Zhao, Yunliu Yang, Juishen Chiao

Department of Microbiology, Shanghai Institute of Plant Physiology, Academia Sinica, 200032 Shanghai, People's Republic of China

Received 25 March 1999; received in revised form 14 June 1999; accepted 9 July 1999; Received by A.M. Campbell.

Abstract

A ~4.8 kb *Kpn*I fragment, from the upstream region of the methylmalonyl-CoA mutase gene (*mutAB*) of rifamycin SV-producing *Amycolatopsis mediterranei*, was cloned and partially sequenced. Codon preference analysis showed three complete ORFs. ORF2 is internal to ORF1, and encodes a polypeptide corresponding to 172 amino acids, whereas ORF1 encodes a polypeptide of 421 amino acids. They were identified as the encoding genes of aspartokinase α - and β -subunits by comparing the amino acid sequences with those in the database. The downstream ORF3, whose start codon was overlapped with the stop codon of both ORF1 and ORF2 by 1 bp, was identified as the aspartate semialdehyde dehydrogenase gene (*asd*), encoding a polypeptide of 346 amino acids. Subclones containing either the *ask* gene or the *asd* gene were constructed, in which the genes could be expressed under *Lac* promoters. Two subclones could transform *E. coli* CGSC 5074 (*ask*-) and *E. coli* X6118 (*asd*-) to prototrophy, supporting the functional assignments. Southern hybridisation indicated that the ~4.8 kb sequenced region represented a continuous segment in the *A. mediterranei* chromosome. It is concluded that *ask* and *asd* genes are present in an operon in *A. mediterranei*, and therefore that organisation of these two genes is the same as in most gram-positive bacteria, such as *Mycobacterium*, *Corynebacterium glutamicum* and *Bacillus subtilis*, but is different from *Streptomyces akimoshiiensis*. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: *Amycolatopsis mediterranei*; Aspartate semialdehyde dehydrogenase; Aspartokinase; Operon

1. Introduction

In most bacteria, amino acids of the aspartate family, such as lysine, methionine and threonine, are produced from aspartate by a series of enzymatic steps. Bacterial biosynthesis of the above amino acids differs from most other pathways of amino acid biosynthesis in that one of the intermediates, diaminopimelate (DAP), has an important metabolic function in its own right as a constituent of the bacterial cell wall peptidoglycan. To ensure that the amino acids produced do not interfere

with cell wall biosynthesis, a complex regulatory system was installed in this pathway. Collectively, these pathways constitute the so-called aspartate pathway, a branched pathway whose common steps are regulated to allow the balanced synthesis of the various end products. Several distinct genomic organisations and a diversity of regulatory mechanisms controlling the metabolic flux through this multibranched pathway have been identified in bacteria.

The enzymes catalyzing the first two steps of this core pathway, aspartokinase (Ask, EC2.7.2.4) and aspartate semialdehyde dehydrogenase (Asd, EC1.2.1.11), are well conserved functionally. In *Escherichia coli* there are three isozymes of aspartokinase, regulated by lysine, methionine and threonine, respectively (Theze et al., 1974). In *Bacillus*, the genes encoding three isozymes of Ask, regulated by diaminopimelic acid (DAP), lysine, or lysine and methionine, have been cloned and sequenced (Chen et al., 1989). Only one Ask has been found in *Pseudomonas*, *Brevibacterium*,

Abbreviations: aa, amino acid; bp, base pairs; kb, kilobase pairs; kDa, kilodalton; M_r , molecular weight; ORF, open reading frame; RBS, ribosome-binding site; DAP, diaminopimelate.

☆ The nucleotide sequence data reported in this paper has been submitted to Genbank under the accession number AF134837.

* Corresponding author. Present address: ISBDD, Suite 212b, 800 East Leigh St., Virginia Biotechnology Park, Richmond, VA 23220, USA.

E-mail address: wzhang@hsc.vcu.edu (W. Zhang)

Mycobacteria, *Thermus* and *Corynebacteria* (Shiio and Miyajima, 1969; Cohen et al., 1969; Cirillo et al., 1994; Cremer et al., 1988; Nishiyama et al., 1995; Jetten et al., 1995).

The *asd* gene has been cloned from many microorganisms, such as *E. coli* (Haziza et al., 1982), *Haemophilus influenzae*, *Salmonella typhimurium* (Galan et al., 1990), *Streptococcus mutans* (Jaguztyn-Krynica et al., 1982; Cardineau and Curtiss, 1987), *Corynebacterium glutamicum* (Kalinowski et al., 1990, 1991), *Saccharomyces cerevisiae* (Thomas and Surdin Kerjan, 1989), *Mycobacteria* (Cirillo et al., 1994), *Leptospira interrogans* (Baril et al., 1992) and *Pseudomonas aeruginosa* (Hoang et al., 1997). The *E. coli asd* gene appears to be regulated principally by lysine and secondarily by methionine and threonine (Haziza et al., 1982). The *ask* and *asd* genes form an operon in gram-positive *Mycobacteria*, *Bacillus* and *Corynebacteria*. The first *asd* gene to be isolated from an antibiotic-producing *Streptomyces* (*Streptomyces akiyoshiensis*) was sequenced recently. The deduced amino acid sequences of the ORFs adjacent to *asd* showed no similarity to any *ask* gene, therefore it seems that streptomycetes differ from other gram-positive bacteria in the organisation of *ask* and *asd* (Le et al., 1996).

In streptomycetes, diaminopimelate (DAP) from the aspartate pathway is an important constituent for their cell wall (Zakharova et al., 1980). Aspartate could be a precursor for both protein amino acids and some secondary metabolites, and it has been shown that regulatory mechanisms acting on aspartate pathways control the flow of metabolic intermediates to some antibiotics (Vining et al., 1990). To date, no aspartokinase gene has been cloned from actinomycetes, and knowledge of the gene structure and regulatory mechanism of the aspartate pathway in actinomycetes is limited. To better understand the aspartate pathway in an antibiotic-producing actinomycete, and its relationship with secondary metabolism, we cloned and sequenced the *ask* and *asd* operon from rifamycin SV-producing *Amiclatopsis*

mediterranei, and expressed them heterologously in *E. coli*.

2. Results and discussion

2.1. Cloning and sequencing of a ~4.8 kb segment from the upstream region of *A. mediterranei mutAB*

We have already determined the nucleotide sequence of ~7.8 kb *KpnI* segment of pCZ8, which contained the methylmalonyl-CoA mutase (*mutAB*) genes and a novel kinase gene (Zhang et al., in preparation). In the present study, a ~4.8 kb *KpnI* fragment, which is ~2 kb upstream of the above ~7.8 kb *KpnI* fragment, was cloned into pUC18, generating pCZ6. ~3 kb *KpnI/PstI* fragment of pCZ6 was sequenced. Three complete ORFs were found in this fragment (Fig. 1). ORF1 starts at nt 220–222 (GTG), which was preceded by a putative streptomycete-like RBS (GGAG) at nt 208–211, and terminates at TGA (nt 1483–1485). By computer-aided analysis, this ORF was found to encode a protein of 421 amino acids ($M_r=44.4$ kDa), with significant sequence similarity to aspartokinases from bacteria (Table 1). Aspartokinase from most gram-positive bacteria are encoded by two in-phase overlapping genes. The high degree of similarity of the *A. mediterranei* aspartokinase (ORF1) with those of *Mycobacteria*, *B. subtilis* and *C. glutamicum* led us to expect a similar gene organisation, indeed, a second ORF (ORF2) was found within ORF1. ORF2 starts at nt 967–969 (GTG), and ends at nt 1483–1485, is preceded by a putative streptomycete-like RBS (GGAG at nt 957–960), and encodes a putative polypeptide of 172 amino acids ($M_r=18.2$ kDa). Both ORF1 and ORF2 use the same stop codon. We therefore tentatively name the two ORFs *askA* and *askB* for the large (α) and small (β) subunits of aspartokinase, respectively. The mean G+C% contents in ORF1 and ORF2 are 68.01% and 66.6%, respectively, which are typical values for actino-

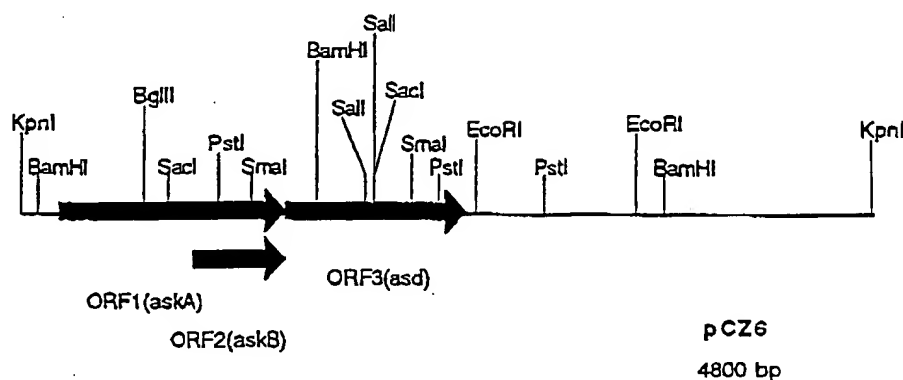


Fig. 1. The restriction map of pCZ6. The relative position and orientation of ORFs are indicated in bold and by arrows.

Table 1
Identities and similarities among some aspartokinases and aspartate semialdehyde dehydrogenases from prokaryotic sources

	Amycolatopsis mediterranei similarity (identity) (%) ^a	
	Aspartokinase	Aspartate semialdehyde dehydrogenase
<i>Mycobacterium smegmatis</i>	88.12 (74.11)	80.70 (68.13)
<i>M. tuberculosis</i>	79.17 (62.99)	81.34 (71.72)
<i>Corynebacterium glutamicum</i>	85.27 (73.16)	76.04 (63.32)
<i>C. florum</i>	85.75 (73.40)	76.33 (63.61)
<i>Pseudomonas aeruginosa</i>	69.00 (46.00)	48.82 (29.29)
<i>Bacillus subtilis</i>	65.11 (44.47)	62.98 (46.15)
<i>Aquifex aeolicus</i>	65.00 (44.00)	66.00 (51.00)
<i>Thermus flavus</i>	70.12 (49.38)	–
<i>T. thermophilus</i>	67.00 (46.00)	–
<i>B. stearothermophilus</i>	68.73 (47.40)	–
<i>Methanococcus jannaschii</i>	59.00 (34.00)	50.00 (35.00)
<i>E. coli</i>	54.06 (28.43) ^b	45.40 (26.71)
<i>Archaeoglobus fulgidus</i>	52.00 (30.00)	47.00 (34.00)
<i>Streptomyces alkydshiensis</i>	–	81.25 (72.62)
<i>Actinobacillus pleuropneumoniae</i>	–	62.54 (46.53)
<i>Streptococcus mutans</i>	–	61.40 (42.25)
<i>Vibrio cholerae</i>	–	61.42 (45.66)
<i>Campylobacter jejuni</i>	–	64.20 (45.68)
<i>Shewanella sp.</i>	–	64.35 (46.52)
<i>Bordetella pertussis</i>	–	47.61 (29.34)
<i>Burkholderia pseudomallei</i>	–	50.00 (32.14)
<i>Haemophilus influenzae</i>	–	47.15 (27.63)
<i>Leptospira interrogans</i>	–	52.81 (33.00)
<i>Salmonella typhimurium</i>	–	48.67 (28.78)
<i>Methanobacterium thermoautotrophicum</i>	–	46.00 (31.00)

^a Values were derived with the GAP program from the University of Wisconsin Genetics Computer Group Package, with a gap weight of 3 and a gap length of 0.1. *Vibrio cholerae* (U25082), *Shewanella sp.* (D49539) (Kato et al., 1997), *Campylobacter jejuni* (X97964) (Pawelec and Jagusztyn-Krynicka, 1996), *Leptospira interrogans* (S92223), *Bordetella pertussis* (X75813), *Actinobacillus pleuropneumoniae* (U51440), *T. thermophilus* (AB013131), *Pseudomonas aeruginosa* ask (AF061757), *Methanococcus jannaschii* (Q57991) (Bult et al., 1996), *Archaeoglobus fulgidus* ask (AE001056) and *asd* (AE000998) (Klenk et al., 1997), *Aquifex aeolicus* (AE000726) (Deekert et al., 1998), *Methanobacterium thermoautotrophicum* (AE000858) (Smith et al., 1997).

^b Lysine-sensitive aspartokinase III of *E. coli* (U00096) (Blattner et al., 1997).

mycete ORFs. The upstream regions of the two ORFs showed no sequences corresponding to *E. coli* σ^{70} -like –10 and –35 hexamers (Strohl, 1992).

In the gram-positive bacteria *Mycobacteria*, *Bacillus* and *Corynebacteria*, *askA* and *askB* are organised in one operon with aspartate semialdehyde dehydrogenase (*asd*), but this is not the case in *S. alkydshiensis*. The *askAB* genes from *A. mediterranei* have high similarity to other cloned *ask* genes, so we also expected to identify the *asd* gene nearby, the *BlasI* analysis showed that downstream ORF3, which overlapped by 1 bp with the *askAB* genes, has 45–71% identity with *asd* genes from gram-positive bacteria, and 28–33% identity with those from gram-negative bacteria. ORF3 starts at nt 1485–1487 (ATG), is preceded by a putative RBS (GGAG) at nt 1472–1475, ends at nt 2523–2525, has a G+C% of 70.34%, and encodes a single polypeptide of 36.4 kDa. According to the *BlasI* results, we tentatively identify ORF3 as the *asd* gene. The *askAB* and *asd* genes possess overlapping stop and start at nt 1485, a device which is thought to lead to translational coupling and which is also found in *A. mediterranei* *mutAB* genes, but not seen

in the *Mycobacteria*, *Bacillus* and *Corynebacteria* *askAB*–*asd* operons. The order of *askAB* and *asd* gene organisation in *A. mediterranei* is the same as *Mycobacteria* and *Corynebacteria*, but is different from that in *Bacillus*, in which the order of the *askAB* and *asd* genes is reversed (Kalinowski et al., 1990).

To prove the *ask* and *asd* genes are located the same way in the chromosome, *A. mediterranei* genomic DNA and pCZ6 were extracted and digested completely with *KpnI* and probed by Southern hybridisation with a P³²-labelled *KpnI* 4.8 kb insert of pCZ6. A clear 4.8 kb band could be identified in the genomic DNA lane, indicating that the cloned 4.8 kb *KpnI* fragment represented the true sequence from *A. mediterranei* (data not shown). These observations suggested that the *askAB* and *asd* genes might be co-transcribed in one mRNA as an operon. The hybridisation results also suggest that there is only one *ask* and one *asd* gene in *A. mediterranei*. Further study is on the way to determine the *ask*–*asd* promoter.

A perfect inverted repeated sequence 12 bp in length capable of forming a stable stem-loop structure, which

	1						70
Cg	MALVVQKYGG	SSLESAERIR	NVAERIVATK	KAGNDVVVVV	SAMGDTTDEL	LELAAAVNPV	PPAREMDMLL
Cf	MALVVQKYGG	SSLESAERIR	NVAERIVATK	KAGNDVVVVV	SAMGDTTDEL	LELAAAVNPV	PPAREMDMLL
Ms	MALVVQKYGG	SSVADAERIR	RVAERIVATK	KAGNDVVVVV	SAMGDTTDEL	LDLARQVSPA	PPPREMDMLL
Am	VALVVQKYGG	SSLESADRIK	RVAERIVATK	KAGNDVVVVV	SAMGDTTDEL	LDLAQQVNPA	PPPERMDMLL
Mt	MALVVQKYGG	SSVADAERIR	RVAERIVATK	KQGNDDVVVV	SAMGDTTDEL	LDLAQQVCFA	PPPREMDMLL
Bs	MGLIVQKFGG	TSVGSVEKIQ	NAANRAIAEK	QKGHQVVVVV	SAMCKSTDEL	VSLAKAISDQ	PSKREMDMLL
Bt	MGIIIVQKFGG	TSVGSIERIQ	HVANRVIEEV	QKCNDDVVVV	SAMGKTDEL	VNLAKQISNH	PSKREMDMLL
Tf	MALVVQKYGG	TSVGDLEIRH	KVAQRIAHYR	EKGHRLAVVV	SAMGHTDEL	IALAKRVNPR	PPPRELDLLT
	*** **	*** **	* *	..	*** **	**	*** **
	71						140
Cg	TAGERISNAL	VAMAIESLGA	EAQSFTGSQA	GVLTERHGN	ARIVDVTGPR	VREALDEGKI	CIVAGFQGVN
Cf	TAGERISNAL	VAMAIESLGA	EAQSFTGSQA	GVLTERHGN	ARIVDVTGPR	VREALDEGKI	CIVAGFQGVN
Ms	TAGERISNAL	VAMAIESLGA	QARSFTGSQA	GVITTTGTHGN	AKIIDVTGPR	LRDALDEGKI	VLVAGFQGVN
Am	TAGERISNLS	VAMAIAAQA	EAWSFTGSQA	GVVTTSVHGN	ARIIDVTGPR	VTEALDQGYI	ALVAGFQGVA
Mt	TAGERISNAL	VAMAIESLGA	HARSFTGSQA	GVITTTGTHGN	AKIIDVTGPR	LQTALEEGRV	VLVAGFQGVN
Bs	ATGEQVTSISL	LSMALQEKGY	DAVSYTGWQA	GIRTEAHGN	ARITDIDTSV	LADQLEKGI	VIVAGFQGMT
Bt	STGEQVTSIAL	LMSLWEKGY	KAVSLTGWQA	GITTEEMHGN	ARIMNIDTTR	IRCLDEGAI	VIVAGFQGMT
Tf	TTGQVSVVAL	LSMQLWAMGI	PAKGFVQHGI	GITTDGRYGD	ARILEVNPAR	IREALDQCFV	AVIAGFQGMT
							*** **
	141						210
Cg	KETRDVTTLG	RGGSDTTAVA	LAAALNADVC	EIYSDVDGVY	TADPRIVPNA	QKLEKLSFEE	MLELAAVGSK
Cf	KETRDVTTLG	RGGSDTTAVA	LAAALNADVC	EIYSDVDGVY	TADPRIVPNA	QKLEKLSFEE	MLELAAVGSK
Ms	QDSKDVTTLG	RCCSDTTAVA	VAAALDADVC	EIYTDVGDIF	TADPRIVPNA	RHLDTVSFEE	MLEMAACGAK
Am	QDTKDITTLG	RGGSDTTAVA	LAAALNADVC	EIYSDVDGVY	TADPRIVPNA	KKLDTVTYEE	MLELAAGSK
Mt	QDTKDITTLG	RGGSDTTAVA	MAAALGADVC	EIYTDVGDIF	SADPRIVRNA	RKLDTVTFEE	MLEMAACGAK
Bs	EDC. EITTLG	RGGSDTTAVA	LAAALKVDKC	DIYTDVPGVF	TTDPRIYKSA	RKLEGISYDE	MLELANLGAG
Bt	ETC. EITTLG	RCCSDTTAVA	LAAALKAEKC	DIYTDVPGVF	TTDPRIYKTA	RKLEISYDE	MLELANLGAG
Tf	PEG. EITTLG	RCCSDTTAVA	LAAALGAKEC	EIYTDTEGVY	TTDPHLIPEA	RKLSVIGYDQ	MLEMAALGAR
	*****	*****	****	.. *	..		*** **
	211						280
Cg	ILVLSRVEYA	RAFNVPLRVR	SSYSNDPGTL	IAGSMEDIPV	E. EAVLTGVA	TDKSEAKVTV	LGISDKPGEA
Cf	ILVLSRVEYA	RAFNVPLRVR	SSYSNDPGTL	IAGSMEDIPV	E. EAVLTGVA	TDKSEAKVTV	LGISDKPGEA
Ms	VLMLRCVEYA	RRYVPIHVR	SSYSOKPCTI	VKGSIEDIPM	E. DAILTGVA	HDRSEAKVTV	VGLPDVPGYA
Am	ILHLRSVEYA	RRYCVPIRVR	SSYSOKPGTT	VTGSIEEIPV	E. QALITGVA	HDRSEAKVTV	TGVDPHTGAA
Mt	VLMLRCVEYA	RRHNPVHVR	SSYSRDPGTV	VVGSIKDVPM	E. DPILTGVA	HDRSEAKVTI	VGLPDIPGYA
Bs	VLHPRAVEFA	KNYQVPLEVR	SSFTETEAGTL	IE...EESM	EQNLIVRGIA	FEDQITRTVI	YGLTSLGLTL
Bt	VLHPRAVEFA	KNYEVPLEVR	SSMENERGTM	VK...EEVSM	EQHLIVRGIA	FEDQVTRTVI	VGLTSLGLTL
Tf	VLHPRAVVYA	KRYGVVHVR	SSFSYNPGTL	VK...EVAM	EMDKAVTGVA	LDLDHAQIGL	IGIPDPGIA
	* ** **	** **	** **	..	*	*	*
	281						350
Cg	AKVFRALA...	DAEINIDM	VLQNVSSVED	GTTDITFTCP	RSDDRAMEI	LKKLQVQGNW	TNVLDDQVQ
Cf	AKVFRALA...	DAEINIDM	VLQNVSSVED	GTTDITFTCP	RADGRAMEI	LKKLQVQGNW	TNVLDDQVD
Ms	AKVFRALA...	EADVNIDM	VLQNIKSID	GKTDITFTCA	RDNQPRAVEK	LSALKSEIGF	SQVLYDDHIG
Am	ARIPRVIA...	DAEIDIDM	VLQNVSSVTS	GRTDITFTLS	KANGAKAVKE	LEKVQAEIGF	ESVLYDDHVG
Mt	AKVFRAVARR	RQHRHGAAE	RLQGRGRQDR	HHLHL...P	QTSQPPWKN	WTRSETRSAS	TQLLYDDHIG
Bs	STIFTTLA...	KRNINVDI	IIQTQAEDKT	G...ISFSVK	TEADADQTVAV	LEEYKDALEF	EKIETESKLA
Bt	ATIPFALA...	NRGINVDI	IIQATNSET	AS...VSFSIR	TEDLPETLQV	L...QALEG	ADVHYESGLA
Tf	AKVFQALA...	ERGIAVDM	IIQGVPGHDP	SRQMAFTVK	KDFAQEALEA	LEPVLAEIGG	EAI...LAPDIA
	351						420
Cg	KVSLVGAGMK	SHPGVTAEFM	EALRDVNVNI	ELISTSEIRI	SVLIREDDLD	AAARALHEQF	QLGGEDEAVV
Cf	KVSLVGAGMK	SHPGVTAEFM	EALRDVNVNI	ELISTSEIRI	SVLIREDDLD	AAARALHEQF	QLGGEDEAVV
Ms	KVSLIGAGMR	SHPGVTATFC	EALAEAGINI	DLISTSEIRI	SVLIKDTELD	KAVSALHEAF	GLGGDEAVV
Am	KVSVVGAGMR	SHPGVTATFC	EALAEAGVNI	EIINTSEIRI	SVLIRDAQLD	DAVRAIHEAF	ELGGDEAVV
Mt	KVSLIGAGMR	SHPGVTATFC	EALAAVGUNI	ELISTSEIOR	SRCC.....	AATPNWTRPW	SRCKMRSGSA
Bs	KVSIIVGSGMV	SNPCVAAZMF	AVLAQKNILI	KMVSTSEIKV	STVVSNDMV	KAVESLHDAF	ELSKHPSAV
Bt	KVSIIVGSGMI	SNPGVAAZMF	EVLADQCIEI	KMVSISEIKI	STVIDEKYV	SAVEELHEAF	GLAEAAAVR
Tf	KVSIIVGVCLA	STPEVPAKMF	QAVASTGANI	EMIATSEVRI	SVIIPAAYAE	AALRAVHQAF	ELDKA.....
	*** **	*** **	*** **	*** **	*** **	*** **	*** **
	421						438
Cg	YAGTGRd...						
Cf	YACTGR...						
Ms	YAGTGR...						
Am	YAGSGR...						
Mt	ATRRPRCTRG	RDGRWACQ					
Bs							
Bt	S...						
Tf							

Fig. 2. Multiple alignment of *A. mediterranei* Ask and some Asks from other sources. The alignment was generated using GCG Pileup. The positions conserved in all Ask proteins are marked by an asterisk. Ms: *Mycobacterium smegmatis* (Z17372), Mt: *M. tuberculosis* (U90239), Cf: *Corynebacterium flaviu* (L16848), Cg: *C. glutamicu* (X57226), Tf: *Thermus flavus* (D37928), Bt: *Bacillus stearothermophilus* (L46351) (Cantoni et al., 1996), Bc: *B. subtilis* (J03294), Am: *A. mediterranei* (AF134837).

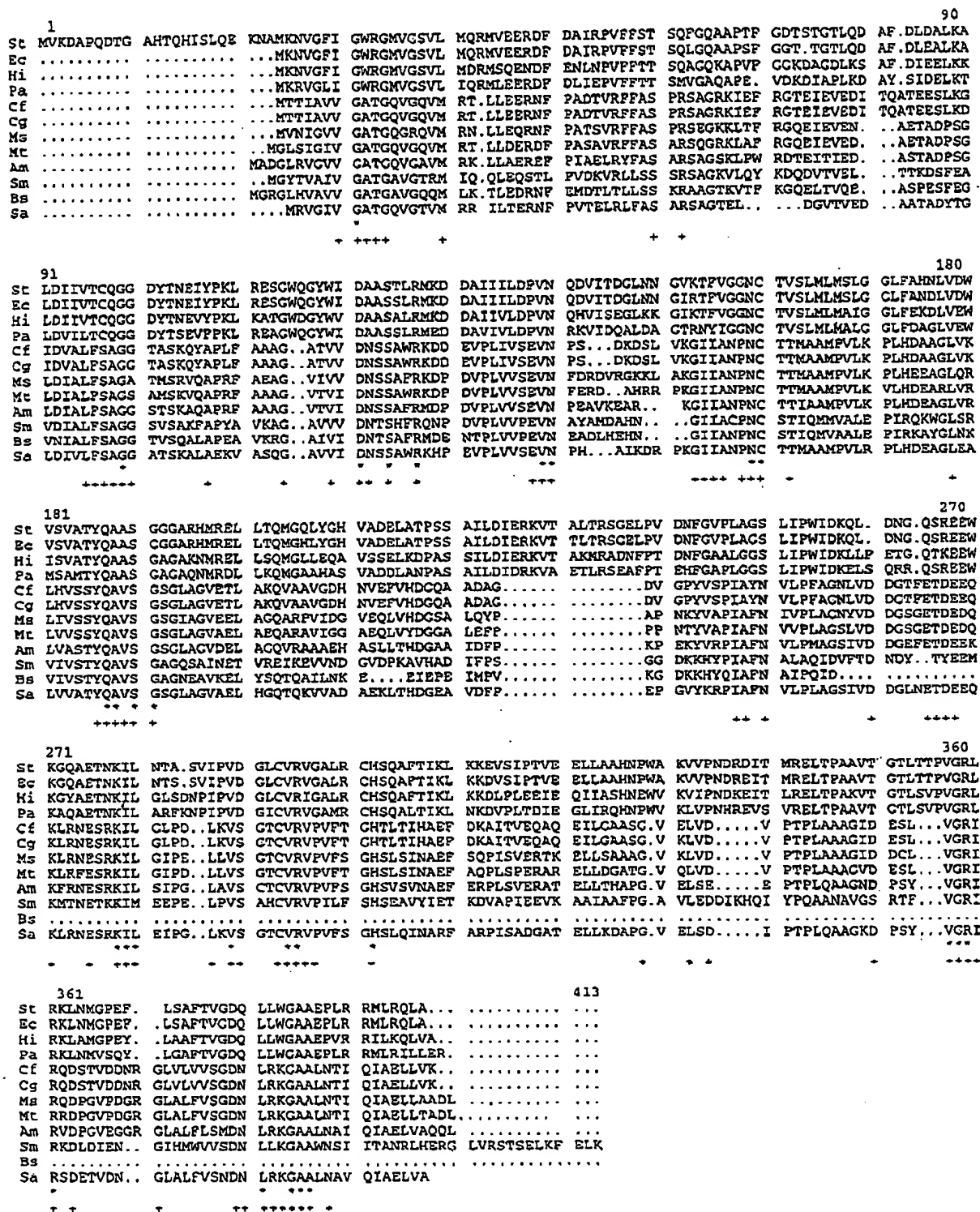


Fig. 3. Multiple alignment of *A. mediterranei* Asd and some Asds from other sources. The alignment was generated using GCG Pileup. The positions conserved in all proteins are marked by an asterisk. The positions conserved only in Asd proteins from gram-positive bacteria are marked +. Ms: *M. smegmatis* (Z17372), Mc: *M. tuberculosis* (U90239), Sa: *Streptomyces akiyoshiensis* (U29446), Cf: *Corynebacterium flavum* (L16848), Cg: *C. glutamicum* (X57226), Ec: *E. coli* (V00262), Pa: *Pseudomonas aeruginosa* (U11055), St: *Salmonella typhimurium* (AF015781), Bs: *B. subtilis* (Z22554), Sm: *S. mutans* (J02667), Hi: *Haemophilus influenzae* (U32747), Am: *A. mediterranei* (AF134837).

IR probably serves as a ρ -independent transcriptional terminator, was found 13 bp downstream of the *asd* gene.

2.2. Analysis of the deduced aa sequence of the *askAB* and *asd* genes

Comparison of the deduced sequence of aspartokinase, encoded by *askAB* of *A. mediterranei*, with the amino acid sequences of nine other aspartokinases derived from gram-positive and -negative bacteria reveals considerable homology between all aspartokinases (Table 1 and Fig. 2). Aspartokinase from *A. mediterranei* shares 44–74% identical residues with aspartokinases from other gram-positive bacteria, and 28.43% identical residues with *E. coli* aspartokinase III. High similarity is found around the amino terminus and a lower level of identity is also seen around aa 250–350. Several possible 'signature' motifs, ⁵VQKXGGXS¹², ³⁸VVXSAMGXTTDELXXLA⁵⁴ and ⁶¹PXXREXDXLT⁶⁹ are found at the amino terminus. The longest conserved motif ¹⁴⁷TTLGRGGSDTTAVAXAAAL¹⁶⁵ is located in the middle of the *askA* gene. The occurrence of these highly conserved elements in aspartokinases from various bacteria with widely different patterns of allosteric regulation, suggested that their roles in structure relate to catalysis rather than to allosteric regulation. Chen et al. (1989) found that mutation of residue 299 of *B. subtilis* aspartokinase changes the allosteric characteristics of aspartokinase, which lies in a region in which the aspartokinase sequences show high divergence. Observations made on proteolytic fragments of *E. coli* aspartokinase revealed that the aspartokinase activity resides in the first 248 residues (Veron et al., 1985), corresponding with the homology pattern listed here. The carboxyl terminus region shows a lower degree of sequence conservation. This region is thought to be involved in the association of subunits. The β -subunit, which has been found unnecessary for aspartokinase activity and may be involved in maintaining the tertiary structure of aspartokinase enzyme complex, also shows lower identity (Cirillo et al., 1994).

The deduced amino acid sequence of *A. mediterranei* aspartate semialdehyde dehydrogenase shows a very high per cent identity with the sequences of the corresponding enzymes from gram-positive bacteria, with highly conserved regions distributed throughout the protein. Although we found that Asd from *A. mediterranei* could substitute for the *E. coli* Asd, the amino acid identity between this protein and Asd enzymes from gram-negative bacteria, including *E. coli*, is much lower (Table 1). A few regions are conserved in all species (Fig. 3). In conjunction with earlier ligand experiments, Biellman et al. (1980) and Haziza et al. (1982) identified the sequence ¹³⁰FVGGNCTVS¹³⁸ as being part of the active site of Asd from *E. coli*. The cysteine residue in

this motif was thought to be alkylated during the catalytic reaction. Fig. 3 showed that the cysteine in this motif (position 160) is conserved in both gram-positive and -negative bacteria.

2.3. Complementation of *askAB* and *asd* mutants of *E. coli*

To confirm the functions of the putative *askAB* and *asd* genes, the ~1.6 kb *Bam*HI and ~1.8 kb *Bgl*II/*Eco*R1 fragments from pCZ6 (Fig. 1) were subcloned into pUC18 and pUC19, respectively, generating plasmids pCZ61 (containing *askAB*) and pCZ62 (containing *asd*). In both cases, the genes could be expressed from the vector *lac* promoter. The two plasmids were used to transform *E. coli* CGSC5074 (an auxotrophic strain mutated in all three *ask* genes) and *E. coli* X6118 (an auxotrophic strain with mutation in the *asd* gene). Asd complementing colonies were observed on LB agar containing 100 μ g/ml ampicillin and 2 mM IPTG in the absence of diaminopimelic acid (DAP), and Ask complementation was observed on M9 medium with 100 μ g/ml ampicillin and 2 mM IPTG. The successful complementation results suggested that ORF1, 2 and ORF3 encode the *askAB* and *asd* genes, respectively.

3. Conclusions

1. The complete aspartokinase and aspartate semialdehyde dehydrogenase genes from rifamycin SV-producing *A. mediterranei* were cloned, sequenced and expressed in *E. coli*. The genes of *askAB* and *asd* encode proteins of 44.4 kDa, 18.2 kDa and 36.4 kDa, respectively. To our knowledge, this is the first cloned aspartokinase from actinomycetes.
2. The evidence presented here indicates that the *askAB* and *asd* genes of *A. mediterranei* are organised in one operon, which is also the case in gram-positive Mycobacteria, Corynebacteria and Bacillus, but is different from *S. akijashiensis*, in which *ask* and *asd* genes are well separated.
3. A few motifs possibly involved in catalytic activity of aspartokinase were found around the amino terminus and middle of aspartokinase. Aspartate semialdehyde dehydrogenase presented much higher variability between gram-positive and -negative bacteria, and no obvious motif could be found on enzymes from all species.

Acknowledgements

This work was supported by a grant from the National Natural Science Foundation of China. We are

grateful to Professor L.C.Vining for the *E. coli* mutants X6118 and CGSC5074, and to Dr. Fiona Steward for suggestions about the manuscript.

References

- Baril, C., Richaud, C., Fournie, E., Baranton, G., Saint Girons, I., 1992. Cloning of *dapD*, *aroD* and *asd* of *Leptospira interrogans* serovar icterohaemorrhagiae and nucleotide sequence of the *asd* gene. *J. Gen. Microbiol.* 138, 47–53.
- Biellman, J.F., Eid, P., Hirth, C., Jornvall, H., 1980. Aspartate-β-semialdehyde dehydrogenase from *Escherichia coli*. *Eur. J. Biochem.* 104, 59–64.
- Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., et al., 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1462.
- Bult, C.J., White, O., Olsen, G.J., Zhou, L., Leischmann, R.D., et al., 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273, 1058–1073.
- Cantoni, R., Labo, M., De Rossi, E., Riccardi, G., 1996. Sequence of the *Bacillus stearothermophilus* gene encoding aspartokinase II. *Gene* 169, 135–136.
- Cardineau, G.A., Curtiss III, R., 1987. Nucleotide sequence of the *asd* gene of *Streptococcus mutans*: Identification of the promoter region and evidence for attenuator-like sequences preceding the structural gene. *J. Biol. Chem.* 262, 3344–3353.
- Chen, N.Y., Zhang, J.J., Paulus, H., 1989. Chromosomal location of the *Bacillus subtilis* aspartokinase II gene and nucleotide sequence of the adjacent genes homologous to *uvrC* and *trx* of *Escherichia coli*. *J. Gen. Microbiol.* 135, 2931–2940.
- Cirillo, J.D., Weisbrod, T.R., Pascopella, L., Bloom, B.R., Jacobs Jr., W.R., 1994. Isolation and characterization of the aspartokinase and aspartate semialdehyde dehydrogenase operon from mycobacteria. *Mol. Microbiol.* 11, 629–639.
- Cohen, G.N., Stanier, R.Y., LeBras, G., 1969. Regulation of the biosynthesis of amino acids of the aspartate family in coliform bacteria and pseudomonads. *J. Bacteriol.* 99, 791–801.
- Cremer, J., Treptow, C., Eggeling, L., Sahm, H., 1988. Regulation of enzymes of lysine biosynthesis in *Corynebacterium glutamicum*. *J. Gen. Microbiol.* 134, 3221–3229.
- Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., et al., 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* 392, 353–358.
- Galan, J.E., Nakayama, K., Curtiss III, R., 1990. Cloning and characterization of the *asd* gene of salmonella-typhimurium use in stable maintenance of recombinant plasmids in salmonella vaccine strains. *Gene* 94, 29–36.
- Haziza, C., Stragier, P., Patte, J.C., 1982. Nucleotide sequence of the *asd* gene of *Escherichia coli*: absence of a typical attenuation signal. *EMBO J.* 1, 379–384.
- Hoang, T.T., Williams, S., Schweizer, H.P., Lam, J.S., 1997. Molecular genetic analysis of the region containing the essential *Pseudomonas aeruginosa* *asd* gene encoding aspartate-β-semialdehyde dehydrogenase. *Microbiology* 143, 899–907.
- Jaguztyn-Krynicka, E.K., Smorawska, M., Curtiss III, R., 1982. Expression of *Streptococcus mutans* aspartate semialdehyde dehydrogenase gene cloned into plasmid pBR322. *J. Gen. Microbiol.* 128, 1135–1145.
- Jetten, M.S., Follett, M.T., Sinskey, A.J., 1995. Effect of different levels of aspartokinase on the lysine production by *Corynebacterium lactofermentum*. *Appl. Microbiol. Biotechnol.* 43, 76–82.
- Kalinowski, J., Bachmann, B., Thierbach, G., Puehler, A., 1990. Aspartokinase genes *lys-c-alpha* and *lys-beta* overlap and are adjacent to the aspartate β-semialdehyde dehydrogenase gene *asd* in *Corynebacterium glutamicum*. *Mol. Gen. Genet.* 224, 317–324.
- Kalinowski, J., Cremer, J., Bachmann, B., Eggeling, L., Sahm, H., Puehler, A., 1991. Genetic and biochemical analysis of the aspartokinase from *Corynebacterium glutamicum*. *Mol. Microbiol.* 5, 1197–1204.
- Kato, C., Smorawska, M., Li, L., Horikoshi, K., 1997. Comparison of the gene expression of aspartate β-D-semialdehyde dehydrogenase at elevated hydrostatic pressure in deep-sea bacteria. *J. Biochem.* 121, 717–723.
- Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E., et al., 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 390, 364–370.
- Le, Y., He, J., Vining, L.C., 1996. *Streptomyces akiyoshiensis* differs from other gram-positive bacteria in the organization of a core biosynthetic pathway gene for aspartate family amino acids. *Microbiology* 142, 791–798.
- Nishiyama, M., Kukimoto, M., Beppu, T., Horinouchi, S., 1995. An operon encoding aspartokinase and purine phosphoribosyltransferase in *Thermus flavus*. *Microbiology* 141, 1211–1219.
- Pawelec, D., Jaguztyn-Krynicka, E., 1996. Complete nucleotide sequence of the *Campylobacter jejuni* T2Dz *asd* gene. *Acta Microbiol. Pol.* 45, 299–304.
- Shio, I., Miyajima, R., 1969. Concerted inhibition and its reversion by end-products of aspartate kinase in *Brevibacterium flavum*. *Biochem.* 65, 849–859.
- Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H.M., Dubois, J., et al., 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* delta H: functional analysis and comparative genomics. *J. Bacteriol.* 179, 7135–7155.
- Strohl, W.R., 1992. Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. *Nucleic Acids Res.* 20, 961–974.
- Theze, J., Margarita, D., Cohen, G.N., Borne, F., Patte, J.C., 1974. Mapping of the structural genes of three aspartokinases and the two homoserine dehydrogenases of *Escherichia coli* K12. *J. Bacteriol.* 117, 133–143.
- Thomas, D., Surdin Kerjan, Y., 1989. Structure of the *hom2* gene of *Saccharomyces cerevisiae* and regulation of its expression. *Mol. Gen. Genet.* 217, 149–154.
- Veron, M., Guillou, Y., Cohen, G., 1985. Isolation of the aspartokinase domain of bifunctional aspartokinase I homoserine dehydrogenase from *E. coli* K12. *FEBS Lett.* 181, 381–384.
- Vining, L.C., Shapiro, S., Madduri, K., Stutard, C., 1990. Biosynthesis and control of β-lactam antibiotics: the early steps in the 'classical' tripeptide pathway. *Biotechnol. Adv.* 8, 159–183.
- Zakharova, O.D., Terekhova, L.N., Agre, N.S., 1980. Actinomycete formation of asporogenic secondary colonies and the properties of their isolated variants. *Mikrobiologiya* 49, 976–978.

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☒ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.